



International Alliance for
Biological Standardization

11th Annual Statistics Workshop: Big Tent Statistics - Conveying the Importance of Statistical Contributions

October 20-23, 2025
Virtual Meeting

Title: Enhancing Yield Investigations through Interpretable Machine Learning

In this work, we present a collaboration between the Statistics & Data Science team and our Commercial Manufacturing/Process Science team to address complex yield improvement challenges. The investigation focused on identifying high-impact factors influencing yield in a high-dimensional, nonlinear process space using commercial manufacturing data.

Two main challenges emerged in applying data-driven approaches to this problem: (1) the need to analyze nonlinear, multivariate batch record data, and (2) the need to engage more efficiently with process scientists throughout the analytical workflow. To address the first challenge, we applied Interpretable Machine Learning (IML) techniques—specifically Shapley value-based model interrogation on pattern recognition algorithms—to uncover and quantify nonlinear relationships between process variables and yield outcomes. This enabled us to move beyond traditional statistical tools and extract actionable insights from complex data structures. To address the second challenge, we developed a structured workflow and digital infrastructure that integrated domain knowledge into the modeling process. An analysis web dashboard was also deployed to automatically analyze selectable tags from new batches, supporting continuous improvement efforts.

Overall, the data analysis step served as a primary tool for hypothesis narrowing, guided variable selection, and continuous feedback—seamlessly integrating into the manufacturing process's continuous improvement workflow..

